CDS

TECHNICAL MEMORANDUM NO. CIT-CDS 95-023
September, 1995

"A Set-Based Methodology for White Noise Modeling"

Fernando Paganini

Control and Dynamical Systems
California Institute of Technology
Pasadena, CA 91125

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **SEP 1995**	2. REPORT TYPE	3. DATES COVERED **00-00-1995 to 00-00-1995**

4. TITLE AND SUBTITLE **A Set-Based Methodology for White Noise Modeling**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **California Institute of Technology,Control and Dynamical Systems,Pasadena,CA,91125**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**34**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

A Set-Based Methodology for White Noise Modeling

Fernando Paganini
Electrical Engineering, M/S 116-81
California Institute of Technology
Pasadena, CA 91125
Email: fernando@hot.caltech.edu

Abstract

This paper provides a new framework for analyzing white noise disturbances in linear systems: rather than the usual stochastic approach, noise signals are described as elements in sets and their effect is analyzed from a worst-case perspective.

The paper studies how these sets must be chosen in order to have adequate properties for system response in the worst-case, statistics consistent with the stochastic point of view, and simple descriptions that allow for tractable worst-case analysis. The methodology is demonstrated by considering its implications in two problems: rejection of white noise signals in the presence of system uncertainty, and worst-case system identification.

1 Introduction

A general feature of mathematical models in engineering science is the presence of modeling errors, which arise due to poorly understood or highly unpredictable phenomena, or from simplifications deliberately introduced for the sake of model tractability. Essentially two approaches are available to assess the consequences of this error: one is to model the uncertainty in terms of a *set* of allowable perturbations and perform worst-case analysis over this set; the other is to assign the additional structure of a probability *measure* to the error, and perform analysis in the average.

Uncertainty is often the dominant issue in models used for control system design. These models involve substantial approximations (linearizations, unmodeled dynamics) and uncertain parameter values, all of which lead to systematic modeling errors for which the only natural characterization is based on sets. Also, the issue of stability provides an incentive to take the worst-case point of view. This has been the strategy of *robust* control theory, which has developed mathematical tools for the evaluation of stability and performance in the worst case over sets of systems. In this theory, the methodology based on sets is also applied to disturbance signals (another source of uncertainty), by modeling them in terms of a ball in some signal space (e.g. $\mathcal{L}_2$, $\mathcal{L}_\infty$), which motivates the $\mathcal{H}_\infty$ or $\mathcal{L}_1$ criteria for worst-case disturbance rejection. The main motivation for these disturbance models is mathematical convenience, since these performance measures can be directly combined with set descriptions of system uncertainty to analyze robust performance (see, e.g. [15]).

This approach for disturbance modeling is pessimistic, however, since it ignores a substantial amount of information about empirical disturbances. It is often the case that these exhibit broad-band spectral characteristics (*white noise*, or some filtered version), especially when they describe the cumulative macroscopic effect of very high dimensional fluctuations at the microscopic level. The statistics of these phenomena have been very accurately modeled by the theory of stochastic processes. The systematic study of the properties of dynamical systems under stochastic noise, pursued by *stochastic* control theory, often leads to tractable results, the most notable being the classical $\mathcal{H}_2$ (LQG) problem. The main limitation to its applicability is that noise is rarely the prevailing source of uncertainty, and the others do not fit easily into a stochastic description ([20] contains some work in this direction).

The robust performance question one would really want to address in many practical cases is the effect of white noise over sets of systems (the "Robust $\mathcal{H}_2$" problem). Many authors (see [25, 7] and references therein) have addressed this problem in terms of a direct combination of the worst-case and stochastic frameworks, and have succeeded in obtaining upper bounds for system performance. At this time, however, this approach is not developed to a competitive level with other performance measures in robust control. In particular, it is difficult to assess the conservatism of these bounds since they involve a combination of worst-case and average case analysis.

Another example of the difficulty of combining these frameworks is the relation between robust control and mainstream system identification (as in [11]), since the latter relies in the stochastic paradigm for noise. Recent efforts in pursuing this unification in the worst-case setting have once again used a pessimistic view of disturbances, resulting in worst-case identification problems with weak consistency properties ([9, 26]) and high computational complexity ([6, 21]).

In this paper we propose a new methodology for white noise modeling, aimed at resolving these difficulties. The starting point is the following question: how does one decide whether a signal can be accurately modeled as a stochastic white noise trajectory? Deciding this from experimental data leads to a statistical hypothesis test on a finite length signal. In other words, one will accept a signal as white noise if it belongs to a certain set. The main idea of our formulation is to take this set as the *definition* of white noise, and carry out the subsequent analysis in a worst-case setting.

For this methodology to be successful, these sets should:

- Exclude non-white signals (e.g. sinusoids) which are responsible for the conservatism of the $\mathcal{H}_\infty$ and $\mathcal{L}_1$ performance measures.
- Include likely instances of white noise. Here stochastic noise will be used as a guidance for the choice of a *typical set*, but not for average case analysis.
- Have simple enough descriptions to allow for tractable worst-case analysis.

The paper is organized as follows: some notation is established in Section 2. In Section 3, the case of signals over a finite horizon is considered, and set descriptions of white noise are given both from the time and the frequency domain points of view. These sets are analyzed in terms of the worst-case system response and in relation to stochastic noise. Section 4 contains the infinite horizon version. In Section 5, the application of this framework both to Robust $\mathcal{H}_2$ analysis and to worst-case system identification is outlined. Space limitations preclude an extensive development of these directions; the objective here is to show the potential of this methodology. The conclusions are given in Section 6, and some proofs are covered in the Appendix.

2 Assumptions and Notation

We will consider discrete time, causal, linear time invariant (LTI) stable systems of the form $H(\lambda) = \sum_{t=0}^{\infty} h(t)\lambda^t$, where λ is the shift operator. Most of the results will be presented for single input/single output (SISO) systems; for the multivariable case see Section 4.3. In the SISO case we will assume that $h(t) \in l_1$; this implies that the summation

$$r_h(\tau) := \sum_{t=0}^{\infty} h(t+\tau)h(t) \tag{1}$$

converges for each τ, defining the autocorrelation sequence of H, and furthermore that $r_h(\tau)$ is itself an l_1 sequence, i.e. $\sum_{\tau=-\infty}^{\infty} |r_h(\tau)| < \infty$. The frequency response (Fourier transform of $h(t) \in l_1$) is denoted by $H(e^{j\omega})$, and is a continuous function of ω. The Fourier transform of $r_h(\tau)$ is the power spectrum $s_h(\omega) := |H(e^{j\omega})|^2$. Also, the $\mathcal{H}_2$ norm of the system is given by

$$\|H\|_2^2 = r_h(0) = \sum_{t=0}^{\infty} h(t)^2 = \frac{1}{2\pi}\int_0^{2\pi} s_h(\omega)d\omega \tag{2}$$

For some of the frequency domain bounds obtained in this paper, we will further assume that $s_h(w)$ is a function of bounded variation (in $BV[0, 2\pi]$). This means (see [22]) that

$$TV(s_h) := \sup_{P=\{w_1,\ldots,w_p\}} \sum_i |s_h(\omega_{i+1}) - s_h(\omega_i)| < \infty \tag{3}$$

where the supremum is over partitions $P = \{w_1, \ldots, w_p\}$ of the interval $[0, 2\pi]$. $TV(s_h)$ is the *total variation* of s_h. The time domain condition $\sum |\tau\ r_h(\tau)| < \infty$ is sufficient for $s_h(w) \in BV[0, 2\pi]$.

3 The Finite Horizon Case

A reasonable starting point for white noise modeling is the case of a scalar valued, finite horizon, discrete time sequence $x(0), \ldots, x(N-1)$ of length N. The infinite horizon version will be considered in Section 4, which also covers the extension to vector-valued signals.

To analyze the response of a system with memory over this finite horizon, some convention must be made on the "past" values of the input signals. The two simplest choices are either to assume the system is initially at rest, or that it is in periodic steady state of period N. We will adopt the latter, since it leads to a more tractable spectral theory: the sequence $x(0), \ldots, x(N-1)$ will be identified with the periodic signal $x(t)$ of period N. This procedure is justified for analyzing stable systems with time constants which are small compared to N, so that the system is not sensitive to long range correlations in the input signals; this will be a standing assumption in this section.

The discrete Fourier transform (DFT) $X(k), k = 0 \cdots N-1$ of the sequence $x(t)$ is defined by the relations

$$X(k) = \sum_{t=0}^{N-1} x(t) e^{-j\frac{2\pi}{N}kt} \quad ; \quad x(t) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kt} \tag{4}$$

The (circular) autocorrelation sequence of x (*correlogram*) is given by

$$r_x(\tau) = \sum_{t=0}^{N-1} x(t+\tau)x(t) \quad \tau = 0 \cdots N-1 \tag{5}$$

and the sequence power spectrum (*periodogram*) by $s_x(k) = |X(k)|^2$, $k = 0 \cdots N-1$.
The sequences $r_x(\tau)$ and $s_x(k)$ form a DFT pair. For an N-periodic signal $x(t)$, we will use as norm the energy over the period, $\|x\|^2 = r_x(0) = \frac{1}{N} \sum_{k=0}^{N-1} s_x(k)$.

The following relations follow immediately from the definitions.

Lemma 1 *Let H be a SISO stable system ($h(t) \in l_1$). If $u(t)$ is an N-periodic input signal to H, and $y = Hu$ is the corresponding steady state (periodic) output, then*

$$(i) \quad r_y(\tau) = \sum_{t=-\infty}^{\infty} r_h(t) r_u(t-\tau) \tag{6}$$

$$(ii) \quad s_y(k) = s_h(\frac{2\pi k}{N}) s_u(k) \tag{7}$$

3.1 White Noise Descriptions in the Time Domain

We wish to characterize white signals among sequences of length N; when faced with the problem of deciding whether an empirical signal is a sample of white noise, a statistician will perform a hypothesis test in terms of some statistic. A common choice (see [2, 11]) is the sample correlogram, which should approximate the expected correlation for white noise (a delta function). In other words a scalar signal is $x(t)$ categorized as white if $r_x(\tau)/r_x(0)$ is small for τ in a certain range (e.g. $1 \leq \tau \leq T$). For example[1], one can choose to specify that the correlogram (normalized to $r_x(0) = 1$), must fall inside a band around zero, of width γ, as depicted in Figure 1.

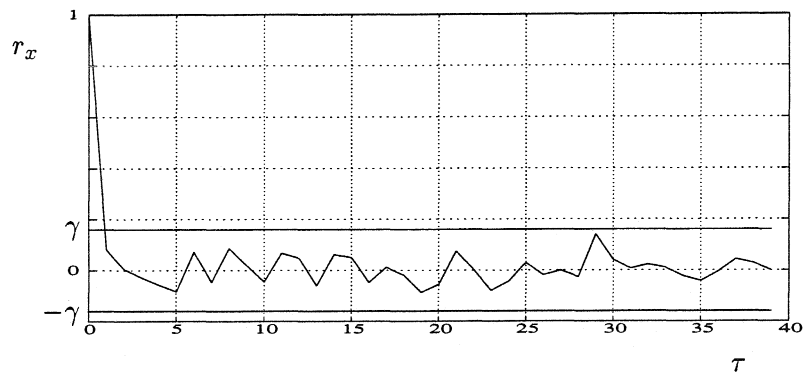

Figure 1: Correlogram of a pseudorandom sequence

From the classical statistical point of view, the choice of γ is associated to a level of significance of the test, which in turn depends on some stochastic model. But regardless of the reasoning behind this choice, ultimately the "whiteness" of the signal is decided in terms of whether it belongs or not to a parametrized set. This motivates the following:

Definition 1 *The set of signals of length N which are white in the time domain sense (accuracy γ, up to lag T) is defined by*

$$W_{N,\gamma,T} := \{x \in \mathbb{R}^n : |r_x(\tau)| \leq \gamma r_x(0), \ \ \tau = 1, \ldots, T\} \tag{8}$$

The response of an LTI system to signals in such sets will now be analyzed from a worst-case perspective. The worst gain of the system under signals in $W_{N,\gamma,T}$ (a seminorm on systems) will be denoted

$$||H||_{W_{N,\gamma,T}} := \sup \left\{ \frac{||y||}{||u||}, \ y = Hu, \ u \in W_{N,\gamma,T}, ||u|| \neq 0 \right\} \tag{9}$$

[1] A common alternative is to bound the sum of the squares of a fixed number of correlogram values; in our context, it is preferable to bound the maximum deviation.

Theorem 2 *Suppose the conditions of Lemma 1 hold, and $u \in W_{N,\gamma,T}$. Then*

$$\left|\frac{r_y(\tau)}{r_u(0)} - r_h(\tau)\right| \leq \gamma \sum_{\substack{t=\tau-T \\ t\neq\tau}}^{\tau+T} |r_h(t)| + \sum_{|t-\tau|>T} |r_h(t)| \tag{10}$$

Furthermore,

$$\left| \|H\|^2_{W_{N,\gamma,T}} - \|H\|_2^2 \right| \leq \gamma \sum_{\substack{t=-T \\ t\neq 0}}^{T} |r_h(t)| + \sum_{|t|>T} |r_h(t)| \tag{11}$$

and for H FIR of length T,

$$\|H\|_2^2 \leq \|H\|^2_{W_{N,\gamma,T}} \leq \|H\|_2^2(1-\gamma) + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| \tag{12}$$

Proof: Equation (10) follows immediately from Lemma 1, and the definition of $W_{N,\gamma,T}$. Applying (10) at $\tau = 0$ gives (11). The upper bound in (12) follows from (11), the lower bound from the fact that the delta function is always a signal in the set $W_{N,\gamma,T}$.

□

Remarks:

1. From inequality (10) we conclude that the autocorrelations of y (up to a constant factor $\|u\|^2$) lie in a band centered at the autocorrelations of the filter. Therefore, such a band is a natural set description for *colored* noise, the output of a linear filter under white noise.

2. It can be shown (see [16]) that if $\gamma < \frac{1}{T}$, then for large enough N the upper bound in (12) is achieved. This is no longer true for large values of γ; for example, if $\gamma = 1$, there are no restrictions on the input signal, and the induced norm can be bounded by the $\mathcal{H}_\infty$ norm of the system which in the FIR case is equal to

$$\sup_{\omega} \left(r_h(0) + 2\sum_{\tau=1}^{T} r_h(\tau)\cos\omega\tau \right)^{\frac{1}{2}}$$

and is in general strictly less than the bound (12). The role of γ in this worst-case approach is to parametrize the freedom allowed in the disturbance signal, and results in a worst-case gain which varies from the $\mathcal{H}_2$ norm for $\gamma = 0$ to the $\mathcal{H}_\infty$ norm for $\gamma = 1$.

Although the choice $\gamma = 0$ would give a clean worst-case theory of white noise rejection, it would mean trading the pessimistic disturbance modeling of $\mathcal{H}_\infty$ for an overly optimistic alternative, since a realistic finite horizon signal will not have exactly zero autocorrelations.

3. In the general case, the parameter T also plays a role, and its adequacy depends on the time constants of the system, as follows from (11). The case $T = N - 1$ is considered below.

There is no absolute answer as to what is a "realistic" white noise signal, but the strongest motivation for these disturbances comes form high dimensional fluctuations (e.g. particle agitation). These have been classically modeled as stochastic processes, but could also be interpreted in the context of deterministic chaos (see [23]). In any event, stochastic noise is known to provide a good model, regardless of whether the probability measure is due to chance or is the ergodic measure of a chaotic system. Therefore, a natural requirement for a realistic white noise set $W_{N,\gamma,T}$ is that it should have large probability for stochastic white signals. In the statistical language, this refers to the level of significance of the hypothesis test for white noise. We will analyze this asymptotically, when the length of N of the data record goes to infinity and γ, T are functions of N.

Theorem 3 *For each N let $x_N = (x(0), \ldots, x(N-1))$ be a vector of independent, identically distributed random variables, with zero mean and finite variance, and $\gamma_N > 0$.*

1. *If T is fixed, and $\gamma_N \sqrt{N} \stackrel{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,T}\right) \stackrel{N\to\infty}{\longrightarrow} 1$.*
2. *If the $x(t)$ are bounded, and $\gamma_N \sqrt{\frac{N}{log(N)}} \stackrel{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,N-1}\right) \stackrel{N\to\infty}{\longrightarrow} 1$.*
3. *If the $x(t)$ are Gaussian, and $\gamma_N \frac{\sqrt{N}}{log(N)^{\frac{3}{2}}} \stackrel{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left(x_N \in W_{N,\gamma,N-1}\right) \stackrel{N\to\infty}{\longrightarrow} 1$.*

Remarks:

- Part 1 of Theorem 3 follows easily from well known results on asymptotic normality of the correlogram: there is substantial averaging between the length N of the time series and the statistic of length T which is employed.

- Parts 2 and 3, with T set to $N-1$, are deeper since there is no averaging: we are imposing constraints of essentially the same dimension as the sample length. These statements are apparently not found in the statistical literature; a proof is given in the Appendix.

The previous theorem has provided a very tight "typical set" for stochastic white noise: we argue that for many purposes, we can now ignore the probability measure and perform worst-case analysis over this set. One such case is disturbance rejection: by choosing $\gamma_N \overset{N\to\infty}{\longrightarrow} 0$ at a sufficiently slow rate, we find that the set $W_{N,\gamma,N-1}$ has asymptotically probability 1 and also $\|H\|_{W_{N,\gamma,N-1}} \overset{N\to\infty}{\longrightarrow} \|H\|_2$. We have therefore reinterpreted the $\mathcal{H}_2$ norm (asymptotically) as the worst-case gain over a typical set, rather than the average gain. Another situation where the probabilistic assumption can be replaced by a typical set is in the context of system identification, as will be discussed in Section 5.2.

Finally, we remark that this approach to modeling based on sets can be applied regardless of any stochastic assumptions on the noise source: what matters is the *statistical* information (which may be obtained directly from empirical correlograms), not the generating mechanism.

The main pending question at this point is whether the chosen sets lend themselves to tractable worst-case analysis. This will be discussed in Section 5.

3.2 Frequency Domain Descriptions

As the name implies, a *white* signal has flat distribution of energy across frequency, which in the finite horizon case would correspond to a flat periodogram (the DFT of a delta-function correlogram). The "raw" periodogram is typically very erratic, however, as demonstrated in Figure 2. This fact has long been recognized (see, e.g. [2, 5]) in the statistical spectral analysis literature; correspondingly, the standard methods for power spectrum estimation are based on smoothing the periodogram, by some form of local averaging that reduces the fluctuations (the variance). This smoothing is most commonly done by convolution of the periodogram with a window function; an abundant literature (see [8]) has studied shapes and properties of these windows.

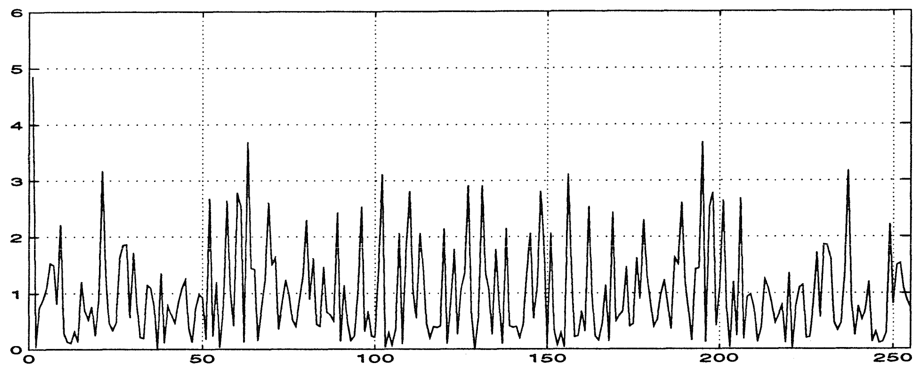

Figure 2: Periodogram of a pseudorandom sequence

In this paper we are interested in defining a set of typical periodograms, which is a hypothesis testing problem. Of course, the image of $W_{N,\gamma,T}$ under the DFT is such a set, but it does not have a simple description in terms of the frequency domain coordinates: the whole purpose of using the frequency domain would be defeated with that description. We will therefore pursue a different characterization for the frequency domain which relies entirely on periodogram properties. One alternative is to specify that a "windowed" version of the periodogram be flat (this was pursued in [16]) but it is preferable to have a test which does not depend on a choice of window.

A very convenient alternative is provided by the Bartlett cumulative periodogram test (see [2, 8]), which consists of accumulating the periodogram and comparing the result to a linear function. Figure 3 contains the result of the accumulation process on the periodogram of Figure 2. As we see, the fluctuations have been smoothed by this integration and the result approximates a linear function in a *uniform* sense; this is the essence of definition which follows.

Definition 2 *The set of white signals in the frequency domain sense, with accuracy η is defined by*

$$\hat{W}_{N,\eta} = \left\{ x \in \mathbb{R}^n : \left| \frac{1}{N} \sum_{k=0}^{m-1} s_x(k) - \frac{m}{N} \|x\|^2 \right| \leq \eta \|x\|^2, \ \ 1 \leq m \leq N \right\} \tag{13}$$

We will now support the frequency domain definition by exhibiting properties which parallel

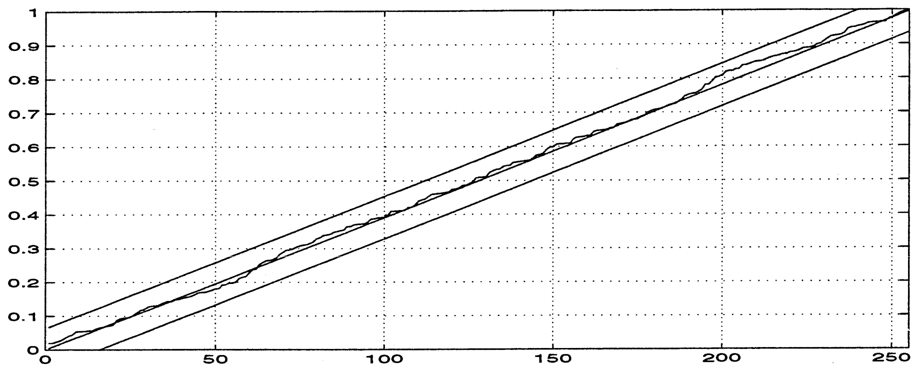

Figure 3: Cumulative periodogram and bounds for $\hat{W}_{N,\eta}$

those in the time domain. The worst-case induced norm of a system H under signals in the set $\hat{W}_{N,\eta}$ will be denoted $\|H\|_{\hat{W}_{N,\eta}}$.

Theorem 4 *Consider a stable LTI system H, with $s_h(\omega) \in BV[0, 2\pi]$. Then*

$$\left| \|H\|_2^2 - \|H\|_{\hat{W}_{N,\eta}}^2 \right| \leq \left(\frac{1}{N} + \eta\right) TV(s_h) \tag{14}$$

Proof:

Fix $u \in \hat{W}_{N,\eta}$, $\|u\|^2 = 1$. Define $\Gamma(k)$ by $\Gamma(0) = 0$, $\Gamma(m) := \frac{1}{N} \sum_{k=0}^{m-1} s_u(k), \;\; 1 \leq m \leq N$. Note that $\Gamma(N) = 1$. Let $y = Hu$, and for simplicity denote $s_h(k)$ in place of $s_h(\frac{2\pi k}{N})$. We have

$$\begin{aligned} \|y\|^2 &= \frac{1}{N} \sum_{k=0}^{N-1} s_h(k) s_u(k) = \sum_{k=0}^{N-1} s_h(k)(\Gamma(k+1) - \Gamma(k)) = \\ &= s_h(N-1) + \sum_{k=1}^{N-1} (s_h(k-1) - s_h(k))\Gamma(k) \end{aligned} \tag{15}$$

Similar calculations show that

$$\frac{1}{N} \sum_{k=0}^{N-1} s_h(k) = s_h(N-1) + \sum_{k=1}^{N-1} (s_h(k-1) - s_h(k)) \frac{k}{N} \tag{16}$$

From (15), (16) we obtain (note that $u \in \hat{W}_{N,\eta}$ implies $|\Gamma(k) - \frac{k}{N}| \leq \eta$)

$$\left| \|y\|^2 - \frac{1}{N} \sum_{k=0}^{N-1} s_h(k) \right| \leq \sum_{k=1}^{N-1} |s_h(k-1) - s_h(k)| \, |\Gamma(k) - \frac{k}{N}| \leq \eta \, TV(s_h) \tag{17}$$

Also, by bounding the difference between the integral $\|H\|_2^2 = \int_0^{2\pi} s_h(w)\frac{d\omega}{2\pi}$ and a step function approximation, it follows that

$$\left| \|H\|_2^2 - \frac{1}{N}\sum_{k=0}^{N-1} s_h(k) \right| \leq \frac{1}{N} \; TV(s_h) \tag{18}$$

which together with (17) leads to (14).

□

In reference to the properties of the set $\hat{W}_{N,\eta}$ in the case of stochastic noise, these have been studied in the statistical literature. We state the following result (see the Appendix):

Theorem 5 *Let $x(0), \ldots, x(N-1), \ldots$ be independent, identically distributed, zero mean Gaussian random variables. If $\eta_N \sqrt{N} \overset{N\to\infty}{\longrightarrow} \infty$, then $\mathcal{P}\left((x(0), \ldots, x(N-1)) \in \hat{W}_{N,\eta}\right) \overset{N\to\infty}{\longrightarrow} 1$.*

These asymptotic properties show that the frequency domain definition is adequate from the point of view of the objectives of this paper: provided $\eta_N \overset{N\to\infty}{\longrightarrow} 0$, $\eta_N\sqrt{N} \overset{N\to\infty}{\longrightarrow} \infty$ the worst case disturbance rejection measure approaches the $\mathcal{H}_2$-norm of the system, while the class of signals contains asymptotically all typical instances of stochastic white noise. Thus the families of time and frequency domain sets have asymptotically the same properties, although they are different for any fixed N.

4 The Infinite Horizon Case

The role of infinite horizon signals in mathematical modeling is that of an abstraction to capture the behavior of signals and systems over a long, but unspecified horizon; the chosen mathematical framework must extend naturally the finite horizon properties and lead to tractable analysis.

Two frameworks arise naturally for the study of deterministic spectral analysis: bounded power signals and bounded energy (l_2) signals.

4.1 Bounded Power Signals

There is a long historical tradition in a non-stochastic theory of white noise, going as far back as Wiener (see [28]), who considered ergodicity properties to build a spectral theory of stationary signals devoid of probability. For disturbance rejection problems, this approach was followed in Zhou et. al. [30], who considered the class of bounded power signals, defined by

$$\mathcal{BP} = \left\{ x(t) : r_x(\tau) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{N} x(t+\tau)x(t) \text{ exists for each } \tau \right\} \tag{19}$$

This class is well motivated since it includes with probability one trajectories of a strictly stationary ergodic random process. Also, similar properties are obtained in the context of deterministic chaos.

The function $r_x(\tau)$ is the autocorrelation of the signal, and the power $\|x\|_P = (r_x(0))^{\frac{1}{2}}$ plays the role of a seminorm (with some restrictions, see below). Also, Bochner's theorem (see [4]) shows that there exists a spectral distribution function $S_x(\omega), \omega \in [0, 2\pi]$ such that $r_x(\tau)$ is recovered from the Stieltjes integral

$$r_x(\tau) = \frac{1}{2\pi} \int_0^{2\pi} e^{j\omega\tau} dS_x(\omega) \tag{20}$$

Equivalently, there exists a positive spectral measure which is the Fourier transform of $r_x(\tau)$; this allows for periodic effects, which correspond to atoms of this measure. It also includes the case of an absolutely continuous spectrum, with the corresponding spectral density $s_x(\omega) = dS_x(\omega)/d\omega$.

We now proceed to give set descriptions of white noise signals in $\mathcal{BP}$, motivated by the finite horizon definitions. In the time domain, define

$$W_{\gamma,T} := \{x \in \mathcal{BP} : |r_x(\tau)| \le \gamma r_x(0) \;\; \tau = 1, \ldots, T\} \tag{21}$$

In the frequency domain, Definition 2 extends by comparing the cumulative spectrum $S_x(\omega)$ with a linear function:

$$\hat{W}_\eta = \left\{ x \in \mathcal{BP} : |S_x(\omega) - \|x\|_P^2 \omega| \le \eta \|x\|_P^2 \;\; \forall \omega \in [0, 2\pi] \right\} \tag{22}$$

One can also consider the ideal white noise set $W_{0,\infty} = \hat{W}_0$ of signals with autocorrelation equal to a delta function, flat spectral density. In fact, this class contains with probability one trajectories of stochastic white noise (see the Appendix):

Proposition 6 *Let $x = (x(0), \ldots, x(t), \ldots)$ be a sequence of independent, identically distributed random variables, with zero mean and finite variance. Then $\mathcal{P}\left(x \in W_{0,\infty}\right) = 1$.*

We now turn to the properties of a stable system with input in $\mathcal{BP}$. To ensure that the output is a $\mathcal{BP}$ signal poses some rather technical issues which we will not address here (the l_2 setting considered later on is more convenient for this). For the moment let us assume, following [30][2] that both input and output are in $\mathcal{BP}$, and satisfy the basic relations

$$(i) \quad r_y(\tau) = \sum_{t=-\infty}^{\infty} r_h(t) r_u(t-\tau) \tag{23}$$

$$(ii) \quad dS_y(\omega) = |H(e^{j\omega})|^2 dS_u(\omega) \tag{24}$$

The worst-case gain in power for signals in the classes $W_{\gamma,T}$ (or $\hat{W}_\eta$) is defined by

$$\|H\|_{W_{\gamma,T}(\hat{W}_\eta)} := \sup\{\|y\|_{\mathcal{P}} : u \in W_{\gamma,T}(\hat{W}_\eta), \|u\|_{\mathcal{P}} = 1\} \tag{25}$$

It follows immediately from (23) that

$$\|H\|_2^2 \leq \|H\|_{W_{\gamma,T}}^2 \leq \|H\|_2^2 + \gamma \sum_{\tau=-T}^{T} |r_h(\tau)| + \sum_{|t|>T} |r_h(t)| \tag{26}$$

For the frequency domain case assume $s_h \in BV[0, 2\pi]$. Consider $u \in \hat{W}_\eta$, $\|u\|_{\mathcal{P}} = 1$; an integration by parts yields

$$\|y\|_{\mathcal{P}}^2 = \frac{1}{2\pi} \int_0^{2\pi} s_h(\omega) dS_u(\omega) = s_h(0) - \frac{1}{2\pi} \int_0^{2\pi} S_u(\omega) ds_h(\omega) \tag{27}$$

Similarly, $\|H\|_2^2 = s_h(0) - \frac{1}{2\pi} \int_0^{2\pi} \omega ds_h(\omega)$ from where

$$\left| \|y\|_{\mathcal{P}}^2 - \|H\|_2^2 \right| \leq \frac{1}{2\pi} \int_0^{2\pi} |S_u(\omega) - w| \, ds_h(\omega) \tag{28}$$

[2][30] states that if $u \in \mathcal{BP}$, $u \in l_\infty$, and the system is exponentially stable, the output is in $\mathcal{BP}$ and (23-24) hold

and therefore

$$\|H\|_2^2 \leq \|H\|_{\hat{W}_\eta}^2 \leq \|H\|_2^2 + \frac{\eta\, TV(s_h)}{2\pi} \tag{29}$$

In particular, the system $\mathcal{H}_2$ norm can be motivated as the gain in power under signals in $W_{0,\infty} = \hat{W}_0$, or equivalently by the limit norms

$$\lim_{\substack{\gamma \to 0 \\ T \to \infty}} \|H\|_{W_{\gamma,T}}\, ; \quad \lim_{\eta \to 0} \|H\|_{\hat{W}_\eta} \tag{30}$$

It is useful to compare this approach with the one used in [30]. The induced gain in power is used there to motivate the $\mathcal{H}_\infty$ system norm; in contrast, the $\mathcal{H}_2$ norm is presented as a "mixed-induced" norm using different seminorms in input and output spaces (power in the output space, and a spectral seminorm in the input, based on the peak spectral density).

In this paper we use power-to-power for both cases: for the $\mathcal{H}_2$ norm, instead of changing the input seminorm, we constrain the inputs to belong to the class of white signals. This seems more direct (worst-case gain over the class of disturbances one expects to see), and allows for a comparison of the $\mathcal{H}_2$ and $\mathcal{H}_\infty$ system norms. The main advantage of our formulation, however, will be made clear in Section 5, where whiteness constraints are incorporated in robustness analysis.

4.2 l_2 Setting

Although the class $\mathcal{BP}$ is conceptually an adequate non-stochastic framework for white noise signals, it is sometimes inconvenient due to its little mathematical structure. In particular, it is not a vector space (not being closed under addition, see [12]), so it is not a seminormed space. For this reason we will now consider white noise descriptions inside l_2, which has the structure of a Hilbert space.

At first, this seems unnatural, since white noise signals are typically considered to be *stationary*, so they will not decay as time goes to infinity. As far as characterizing system response to signals with flat spectrum, however, the response to l_2 signals is just as representative as the response to bounded power signals: the "behavior at ∞" should not be the determining factor in any sensible engineering model. For example, the response of an LTI system to a bounded power signal is approximately the same as the response to a very long truncation.

Actually, the same considerations apply to standard $\mathcal{H}_\infty$ theory. While the $\mathcal{H}_\infty$ norm is most naturally motivated [30] by the gain in power for bounded power inputs, since this class includes sinusoids, most technical results on $\mathcal{H}_\infty$ are obtained by using l_2 as a signal space, which does not contain these signals, but contains signals of arbitrarily narrow bandwidth.

For l_2 (square-integrable) sequences, the autocorrelation is defined by $r_x(\tau) = \langle x, \lambda^\tau x\rangle$. The corresponding spectral measure is absolutely continuous, with spectral density $s_x = \frac{dS_x}{d\omega} = |X(e^{j\omega})|^2$, where $X(e^{j\omega})$ is the Fourier transform of $x(t)$.

The sets $W_{\gamma,T}$ and $\hat{W}_\eta$ over l_2 can then be defined as in (21) and (22); the same properties hold for system gain, where the signal norm is now taken to be the l_2 norm.

4.3 Multivariable Extension

This section outlines how the previous methodology can be extended to deal with vector valued white noise signals. We will only consider the case of infinite horizon l_2 signals, which demonstrates all the necessary extensions; the same ideas could be applied in a finite horizon setting.

For vector-valued signals $x(t) \in l_2(\mathbb{R}^n)$, the matrix autocorrelation (prime denotes transpose, * denotes conjugate transpose) is given by

$$R_x(\tau) = \sum_{t=-\infty}^{\infty} x(t+\tau)x'(t) \tag{31}$$

Once again, a spectral (matrix) distribution function $S_x(\omega)$ is defined, verifying a matrix version of (20). In this l_2 case $\frac{dS_x(\omega)}{d\omega} = s_x(\omega) = X(e^{j\omega})X^*(e^{j\omega})$, where the column vector $X(e^{j\omega})$ is the Fourier transform of $x(t)$. The 2-norm of the signal verifies

$$\|x\|_2^2 = trace(R_x(0)) = \frac{1}{2\pi}\int_0^{2\pi} trace(s_x(e^{j\omega}))d\omega \tag{32}$$

Consider a stable, discrete time linear time invariant system with in general n inputs and p outputs, $H(\lambda) = \sum_{t=0}^{\infty} h(t)\lambda^t$, with frequency response $H(e^{j\omega})$.

Defining $R_H(\tau) = \sum_{t=0}^{\infty} h(t+\tau)h'(t)$ and $s_H(\omega) = H(e^{j\omega})H^*(e^{j\omega})$, the $\mathcal{H}_2$ norm of H satisfies (32). If $u(t) \in l_2(\mathbb{R}^n)$, $y(t) \in l_2(\mathbb{R}^p)$ are respectively, the input and output to H, then:

$$R_y(\tau) = \sum_{t=0}^{\infty}\sum_{s=0}^{\infty} h(s)R_u(\tau+t-s)h'(t) \tag{33}$$

$$s_y(\omega) = H(e^{j\omega})s_x(\omega)H(e^{j\omega})^* \tag{34}$$

Now we give set descriptions of vector valued white noise. For a matrix A denote $\|A\|_\infty = \max_{i,j}|A_{i,j}|$, and define the following set in terms of time domain constraints

$$W^n_{\gamma,T} := \left\{ x(t) \in l_2(\mathbb{R}^n) : \left\| R_x(\tau) - \delta(\tau)\frac{\|x\|_2^2}{n} I \right\|_\infty \leq \gamma \|x\|_2^2 , \ |\tau| \leq T \right\} \tag{35}$$

In (35) we impose low autocorrelation for $1 \leq \tau \leq T$, and also low "spatial" correlation between the components of the vector signal. The choice of the matrix norm in (35) is somewhat arbitrary; the previous choice has the advantage of imposing quadratic signal constraints (see Section 5).

For a frequency domain characterization, define

$$\hat{W}^n_\eta := \left\{ x \in l_2(R^n) : \left\| S_x(w) - \frac{\omega}{n}\|x\|^2 I_m \right\|_\infty < \eta \|x\|^2 , \ \omega \in [0, 2\pi] \right\} \tag{36}$$

Defining $\|H\|_{W^n_{\gamma,T}}$, $\|H\|_{\hat{W}^n_\eta}$ as usual, bounds similar to (26), (29) can be obtained from (33), (34), leading to

$$\lim_{\substack{\gamma\to 0\\ T\to\infty}} \|H\|_{W^n_{\gamma,T}} = \lim_{\eta\to 0} \|H\|_{\hat{W}^n_\eta} = \frac{1}{\sqrt{n}}\|H\|_2 ; \tag{37}$$

Remarks:

- The factor $\frac{1}{\sqrt{n}}$ arises from the use of the same norm in input and output space. It can be also motivated for stochastic noise: if the input has covariance matrix I, the expected input power is $\sqrt{n}$, and the expected output power is $\|H\|_2$.

- In l_2^n space there are no ideally white signals ($R_x(\tau) = \delta(\tau)I$, or $S_x(w) = wI$), since this would imply $s_x(\omega) = I$, and it is a rank one matrix for each ω. "Pure" white multivariable noise appears only in the bounded power approach. For $\gamma > 0$, $\eta > 0$, however, the l_2^n sets $W_{\gamma,\infty}$ and W_η are non-trivial, giving arbitrary approximations to white noise which can be used via (37) to motivate the $\mathcal{H}_2$ norm within the l_2 framework.

5 Worst-Case Analysis over White Noise Sets

The previous sections have provided set descriptions of white noise signals aimed at worst-case analysis, and have shown that this procedure is sound and gives results which are consistent with the alternative stochastic setting.

We will now show that this approach leads to tractable worst-case analysis by showing applications of this framework to two different problems mentioned in the introduction: robust $\mathcal{H}_2$ performance analysis and worst-case system identification. We will not attempt to present a full description of these directions in the limited space available here; they have been developed elsewhere [18, 17, 19, 27], and journal versions are in preparation. In this paper our objective is to provide enough evidence that this methodology has useful implications.

5.1 Robust $\mathcal{H}_2$ Analysis

A problem which has received substantial attention (e.g. [25, 7]) is that of obtaining robust performance guarantees for a set of systems subject to white noise disturbances.

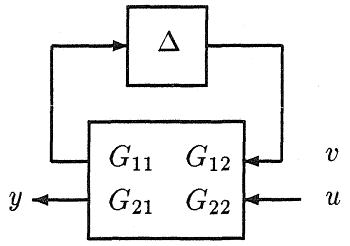

Figure 4: Uncertain system

In the system of Figure 4, G is a known (nominal) map which is assumed to be an LTI system. The perturbation Δ represents the system uncertainty, which is assumed to have block diagonal structure, of the form $\Delta = diag[\Delta_1, \ldots, \Delta_F]$, (each Δ_i is of size q_i) and is normalized to a set $\mathbf{B}_{\boldsymbol{\Delta}} := \{\Delta : ||\Delta|| \leq 1\}$. For background and motivation for this setup, see [15].

The objective is to analyze rejection properties of the system to a white noise disturbance applied in u, in the worst-case over $\Delta \in \mathbf{B}_{\mathbf{\Delta}}$. If the perturbation Δ is assumed to be LTI, this corresponds to finding the worst-case $\mathcal{H}_2$ norm of the closed loop transfer function from u to y; we will mostly deal with linear time-varying (LTV) Δ's here, but still refer to robust $\mathcal{H}_2$ performance with some abuse of terminology.

Set descriptions will be applied to describe the white noise disturbance u; as argued in Section 4.2, it is sufficient to consider the sets $W_{\gamma,T}$ or $\hat{W}_\eta$ inside l_2 space. The robust performance analysis problem (e.g., for $W_{\gamma,T}$) is therefore to compute

$$\sup_{\substack{\Delta \in \mathbf{B}_{\mathbf{\Delta}} \\ u \in W_{\gamma,T}, \|u\|_2 \leq 1}} \|y\|_2 \tag{38}$$

Before addressing this problem we review how this question can be handled when there is no constraint on u.

5.1.1 Background on Robust $\mathcal{H}_\infty$ Analysis

The robust performance question most commonly treated in the literature deals with $\mathcal{H}_\infty$ performance, which refers to the worst-case gain of the system as an operator on l_2. For the system in Figure 4, for the case where Δ is a structured, otherwise arbitrary linear operator on l_2, the following necessary and sufficient condition for has been obtained [24, 14]:

$$\sup_{\substack{\Delta \in \mathbf{B}_{\mathbf{\Delta}} \\ \|u\|_2 \leq 1}} \|y\|_2 < 1 \iff \exists X \in \mathbb{X} : \; G(e^{j\omega})^* \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} G(e^{j\omega}) - \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} < 0 \tag{39}$$

In (39), $\mathbb{X}$ is the set of positive scaling matrices of the form $X = diag[x_1 I_{q_1}, \ldots, x_F I_{q_F}]$. Since $\mathbb{X}$ is convex and so is the condition (39), this test has tractable computational properties (see [3, 15]).

We now briefly explain how this result can be obtained from the Integral Quadratic Constraint (IQC) formulation, which originated in the work of Yakubovich [29], and has recently been applied to this $\mathcal{H}_\infty$ problem by Megretski [14].

Let z be the vector of all the inputs to system G, $z = (z_1, \ldots, z_F, z_{F+1})$, where $(z_1, \ldots, z_F)$ is the partition of the signal v in terms of the blocks of Δ, and $z_{F+1} = u$. Analogously $(Gz)_i, i = 1, \ldots, F+1$ denote the partitions of the output of G.

Now define the following scalar valued quadratic functions of $z \in l_2$,

$$\sigma_i(z) = \|(Gz)_i\|^2 - \|z_i\|^2, \quad i = 1, \ldots, F+1 \tag{40}$$

The main observation is that if the $\sigma_i(z)$, $i = 1, \ldots, F+1$ are all non-negative for a certain $z \neq 0$, then G expands this signal in every channel, and therefore contractive LTV perturbations $\Delta_1, \ldots, \Delta_F$ exist such that the closed loop is expansive at z, violating robust $\mathcal{H}_\infty$ performance.

Robust performance is thus converted to a condition on the sign on a finite number $\sigma_1, \ldots, \sigma_{F+1}$ of quadratic forms on l_2. (39) now follows from the application of the following result by Megretski and Treil [14]: Given $\sigma_1, \ldots, \sigma_{F+1}$, where each $\sigma_i : l_2 \rightarrow \mathbb{R}$ is a shift invariant quadratic form in l_2, the following are equivalent:

1. There does not exist $z \in l_2$ such that $\sigma_i > 0 \;\; i = 1, \ldots, F+1$.

2. There exist $x_i \geq 0$, $i = 1, \ldots, F+1$, not all zero such that $x_1\sigma_1 + \ldots + x_{F+1}\sigma_{F+1} \leq 0$

Note that the only non-trivial direction is 1$\Rightarrow$2; this is called "S-procedure losslessness". Applied to (40), condition 2 then leads to (39), with $X = diag[x_1 I, \ldots, x_F I]/x_{F+1}$. Some refinements of these arguments are needed (see [14]) to obtain strict inequalities, and $x_i > 0$.

5.1.2 Robust Performance Analysis over the Signal Set $W_{\gamma,T}$

We now show the robust performance problem remains tractable when the disturbance u is constrained to vary in the white noise set $W_{\gamma,T}$. We consider for simplicity the case of scalar noise, similar arguments apply to the multivariable case. The main observation is that $W_{\gamma,T}$ is described by a finite number of constraints

$$\gamma r_u(0) \pm r_u(\tau) \geq 0 \quad \tau = 1, \ldots, T \tag{41}$$

which are quadratic on the signal u. In other words, they are IQCs (this was already suggested in [13]) corresponding to the quadratic forms $\sigma_\tau^+(u) = \gamma r_u(0) + r_u(\tau)$, $\sigma_\tau^-(u) = \gamma r_u(0) - r_u(\tau)$.

Using the same arguments as in the case of $\mathcal{H}_\infty$, robust performance analysis over $W_{\gamma,T}$ reduces to the question of whether there exist signals $z \in l_2$ verifying simultaneously

$$\begin{aligned} \sigma_i(z) &\geq 0, \;\; i = 1, \ldots, n+1 \qquad (42) \\ \sigma_\tau^\pm(z_{n+1}) &\geq 0, \;\; \tau = 1, \ldots, T \qquad (43) \end{aligned}$$

We are therefore once again in a position to apply the losslessness theorem cited above, which will imply the existence of non-negative scalings x_i, $x_\tau^\pm$ satisfying

$$\sum_{i=1}^{n+1} x_i \sigma_i + \sum_{\tau=1}^{T} x_\tau^+ \sigma_\tau^+ + \sum_{\tau=1}^{T} x_\tau^- \sigma_\tau^- \leq 0 \qquad (44)$$

The previous condition is once again convex in the scaling parameters x_i, $x_\tau^\pm$, which suggests that computational methods similar to those for robust $\mathcal{H}_\infty$ performance analysis should result. These issues are further studied in [17, 18], where the theory is developed from a different (though equivalent) point of view. The idea in [17, 18] is that IQCs can be represented in *implicit* form, in terms of uncertain equations. Representing the system also in implicit form reduces the problem to a standard form of implicit analysis question which is studied in detail in [17, 18]. In particular, state-space methods are available to compute the convex condition (44).

An important remark is that the losslessness result of [14] only applies to a *finite* number of IQCs, which requires $T < \infty$. As $\gamma \to 0, T \to \infty$ we will approach the robust $\mathcal{H}_2$ performance measure, but convergence in T could be slow, leading to large computations. The situation is better for the frequency domain sets, as is shown next:

5.1.3 Robust $\mathcal{H}_2$ Analysis in the Frequency Domain.

Robustness analysis with white noise signals over the set $\hat{W}_\eta$ appears to be a more complicated problem since $\hat{W}_\eta$ is not described by a finite number of scalar constraints: although the integrated spectrum $S_u(\omega)$ depends quadratically on u, and is shift invariant, it takes values on a function space, namely the space of continuous functions in $[0, 2\pi]$. The constraint

$$|\frac{1}{\|u\|_2^2} S_u(\omega) - \omega| \leq \eta \quad \forall \omega \in [0, 2\pi] \tag{45}$$

imposed on S_u in the definition of $\hat{W}_\eta$ is also infinite dimensional in nature.

It is shown in [19], however, that this approach leads to a very compact solution to the robust $\mathcal{H}_2$ analysis problem. Although it remains infinite dimensional, its form lends itself to simple finite dimensional approximations. The following is the necessary and sufficient condition for Robust $\mathcal{H}_2$ performance (i.e. robust performance over $\hat{W}_\eta$ for small enough η) for the system of Figure 4:

There exists $X \in \mathbb{X}, X > 0$, and a matrix function $\Phi(\omega) \in \mathbb{C}^{m \times m}$, $\Phi = \Phi^*$, such that

$$(i) \quad G(e^{j\omega})^* \begin{bmatrix} X & 0 \\ 0 & I \end{bmatrix} G(e^{j\omega}) - \begin{bmatrix} X & 0 \\ 0 & \frac{1}{n} I + \Phi(\omega) \end{bmatrix} < 0 \tag{46}$$

$$(ii) \quad \int_0^{2\pi} trace(\Phi(\omega)) \frac{d\omega}{2\pi} < 0 \tag{47}$$

The previous condition is stated for the multivariable noise case (n is the dimension of the noise). We see that (46) is very similar to condition (39) for Robust $\mathcal{H}_\infty$ performance; the only addition is the incorporation of the function $\Phi(\omega)$, which plays the role of an infinite dimensional "multiplier" corresponding to the constraints defining $\hat{W}_\eta$. Heuristically, for $n = 1$, $\Phi(\omega)$ allows for the gain to be larger than 1 at some frequency, provided that it is compensated at some other frequency by keeping the total effect $\int \Phi(\omega) d\omega$ negative; this imposes in effect an average over frequency performance which corresponds to the $\mathcal{H}_2$ norm.

We note that the computational properties of this test are also of a similar nature to those of (39). Consequently, analyzing robust $\mathcal{H}_2$ performance is essentially no harder than analyzing robust $\mathcal{H}_\infty$ performance; this is a strong result which is shows the benefits of modeling uncertainty and disturbances in a consistent framework.

A proof of this result is given in [19], where various extensions are also considered, in particular to the case of LTI uncertainty, involving frequency-dependent scaling matrices X.

5.2 Worst-Case System Identification

The classical literature on system identification (see [11] and references therein) characterizes model errors as due to stochastic noise; system identification in this setting is a special case of an estimation problem in statistical inference. From this perspective, the main requirement for an identification scheme is that if the true system is in the model class, the estimates are *consistent*, i.e. they converge to the true values in a stochastic sense, as the length of the experiment goes to infinity.

In contrast, robust control theory has relied on error models based on sets, e.g. a ball of systems in some norm. The desire to make identification and robust control more compatible has stimulated a research direction (see e.g. [9, 26, 6, 21]) which treats the system identification problem from a worst-case point of view, and seeks "hard" bounds on the identification error. In this formulation noise plays the role of an adversary; if, as is standard in robust control, it is allowed to vary over a large class (e.g. a ball in l_∞), then consistency of the estimates can no longer be ensured.

We now discuss these issues in the simple situation of a SISO model structure

$$y = h * u + d \tag{48}$$

where $h = (h(0), \ldots, h(T-1))$ is FIR, and d is noise. Given data for y, u of length N, the problem is to estimate the system h. The equations in (48) can also be written in matrix form as $y = \mathcal{U}h + d$, where y, h, d are column vectors and $\mathcal{U}$ denotes the $N \times T$ Toeplitz matrix with first column u. The 2-norm will be used for signals here; the input is normalized to $\|u\|_2^2 = N$. To simplify the analysis, assume that the experiment was started at time $-(T-1)$, with values of u which are N-periodic.

In the classical theory, d is assumed to be stochastic white noise: IID random variables, with zero mean, variance σ^2. In this linear regression problem the minimum variance estimate for h is given by the least squares solution

$$\hat{h} = (\mathcal{U}^*\mathcal{U})^{-1}\mathcal{U}^*y \tag{49}$$

where invertibility of $\mathcal{U}^*\mathcal{U}$ (*persistence of excitation*) is assumed. The estimator (49) is unbiased, and its covariance matrix $\sigma^2(\mathcal{U}^*\mathcal{U})^{-1}$ will converge to zero as $N \to \infty$, under stationarity assumptions in u. This implies that in the stochastic sense, the estimator will be consistent.

For worst-case identification, we first follow the usual approach which is to only restrict d to be bounded in norm; suppose $\|d\|_2^2 \leq \delta^2 N$ (noise to signal ratio δ). Since there is a linear relation (48) between h and d, the set of h values compatible with the data and the constraint $\|d\|_2^2 \leq \delta^2 N$ will be an ellipsoid. It follows that if one wishes to minimize the maximum error in the 2-norm in h, the optimal choice is the center of the ellipsoid, which once again corresponds to the least squares solution (49). Assuming now for simplicity that u is purely white (i.e. $\mathcal{U}^*\mathcal{U} = NI$, this is also the optimal choice) the worst-case estimation error

$$\|\hat{h} - h\|_2 = \left\|(\mathcal{U}^*\mathcal{U})^{-1}\mathcal{U}^* d\right\|_2 = \frac{1}{N}\|\mathcal{U}^* d\|_2 \tag{50}$$

has a value of δ, corresponding, for example, to $d = \delta u$.

We therefore find that although both points of view lead in this case to the same optimal estimate, they attach to it a different interpretation. In particular, consistency is lost in the worst-case setting: the estimation error cannot be made smaller than δ, no matter how long the data record is. The same was found in [9, 26] for other system norms. The reason for this pessimistic interpretation is that the noise, which plays an adversarial role, is allowed to vary in a class where it can "conspire" to have a high correlation with the input. This suggests that the desirable consistency interpretation can be recovered if the disturbance is constrained in the style of this paper to have low cross correlation with u.

One way of doing this was studied recently by Venkatesh and Dahleh [27]: the input is chosen to be periodic of period T (this allows for persistence of excitation of order T), and the disturbance d is restricted to the set $W_{N,\gamma,N-1}$[3]. The main observation from [27] is that in this case (assuming N is a multiple of T)

$$\|\mathcal{U}^* d\|_2^2 = \sum_{\tau=0}^{N-1} r_d(\tau) r_u^T(\tau) \tag{51}$$

[3] In [27] a variation of this set is used; it leads nevertheless to similar bounds as those given here.

where $r_d(\tau)$ is the correlogram of d (length N) and $r_u^T(\tau)$ is the correlogram for u of length T, repeated periodically. For a purely white u, we would have

$$r_u^T(\tau) = \begin{cases} T \ \text{ for } \tau = kT,\ k = 0, \ldots, \frac{N}{T} - 1 \\ 0 \ \ \tau \neq kT,\ \ 1 \leq \tau \leq N-1 \end{cases} \tag{52}$$

Imposing that $d \in W_{N,\gamma,N-1}$, (50), (51) and (52) give

$$\|\hat{h} - h\|_2^2 \leq \frac{T}{N^2}\left(r_d(0) + \sum_{k=1}^{N/T-1} r_d(kT)\right) \leq \frac{T}{N^2}\delta^2 N\left(1 + \gamma(\frac{N}{T} - 1)\right) \leq \delta^2\left(\gamma + \frac{T}{N}\right) \tag{53}$$

We now consider another way to constrain the identification problem, which is to directly impose low correlation between u and d. For example, we can impose that (u, d) is a white signal in the multivariable sense. More precisely, that $(\delta u, d)$ (scaling both components to the same size) is in the set $W_{N,\gamma,T}^2$. This set is the finite horizon version of the set $W_{\gamma,T}^2$ of (35) and in particular imposes the cross correlation constraints

$$\delta|\langle\lambda^\tau u, d\rangle| = |\langle\lambda^\tau \delta u, d\rangle| \leq \gamma(\|\delta u\|^2 + \|d\|^2) = 2\gamma\delta^2 N \quad \tau = 0, \ldots, T \tag{54}$$

Since the elements of $\mathcal{U}^* d$ are $\langle\lambda^\tau u, d\rangle$, $\tau = 0, \ldots, T-1$, these bounds can be applied to (50) giving

$$\|\hat{h} - h\|_2 = \frac{1}{N}\|\mathcal{U}^* d\|_2 \leq \frac{1}{N} 2\gamma\delta N\sqrt{T} = 2\gamma\delta\sqrt{T} \tag{55}$$

In both cases ((53) and (55)) if $\gamma_N \to 0$ as $N \to \infty$, we obtain the consistency property $\hat{h}_N \overset{N\to\infty}{\longrightarrow} h$.

By choosing an appropriate decay rate for γ (e.g. $\gamma = \frac{1}{N^\alpha}, \alpha < \frac{1}{2}$), the chosen disturbance set has high probability from the stochastic viewpoint (Theorem 3 applies to the case $d \in W_{N,\gamma,N-1}$; a similar argument can be used for the multivariable case, or applied to the constraints (54) alone). Therefore, our class of disturbances is still rich enough to accommodate classical identification. In addition, the errors in (53) and (55) will decay to zero in polynomial time (in contrast to the complexity results of [6, 21]).

As those in [26, 6, 21], these results for FIR identification are mainly of conceptual value, and contribute to understand the properties of the identification problem from a worst-case perspective.

However they provide important practical guidelines as to how a more general identification problem should be posed when worst-case guarantees are sought (e.g. identification involving noise and set descriptions of unmodeled dynamics). To avoid conservatism the disturbance must be constrained explicitly, and correlation constraints are an adequate tool for this. These more general problems are currently under investigation.

6 Conclusion

As a field of engineering science, control theory has a broad interaction with mathematics, drawing on tools from various disciplines, such as dynamical systems, algebra, functional analysis and probability. While these provide a variety of viewpoints which is an asset of this field, it is sometimes difficult to combine the positive features of the different frameworks. In this paper we have succeeded in addressing one such situation, providing a meeting point between the functional analytic and stochastic points of view.

Of course, many problems will not yield to this kind of unification. In particular, not all aspects of a stochastic description can be captured by worst-case analysis over a typical set. Nevertheless, we feel that there is potential for further applications of this line of thinking in various engineering problems, which naturally call for a combination of "hard bounds" and probabilistic models.

Appendix

This section contains proofs and supplementary material for the stochastic results.

Proof of Theorem 3:

Part 1: For the case of a fixed time lag τ, the distribution of the autocorrelation $r_x(\tau)$ has been extensively studied in the statistical literature [2, 1]; exact expressions for the distribution of $r_x(\tau)/r_x(0)$ when $x(t)$ is Gaussian are obtained in [1], and it follows that $\sqrt{N}\frac{r_x(\tau)}{r_x(0)}$ is asymptotically normal $\mathcal{N}(0,1)$. Since $\gamma\sqrt{N} \to \infty$, and T is fixed,

$$\mathcal{P}\left(x \notin W_{N,\gamma,T}\right) \leq \sum_{\tau=1}^{T} \mathcal{P}\left(\left|\sqrt{N}\frac{r_x(\tau)}{r_x(0)}\right| > \gamma\sqrt{N}\right) \overset{N\to\infty}{\longrightarrow} 0 \tag{56}$$

□

In parts 2, 3 of the theorem, the number of correlation constraints grows with the sample size, and the argument with the normal approximation cannot be used: even though each $r_x(\tau)$ for fixed τ is asymptotically normal, the joint distribution of $(r_x(1), \ldots r_x(N-1))$ is defined on a space of increasing dimension, where no global averaging occurs. Our proof relies on a Hoeffding inequality for sums of bounded random variables, [10]:

Theorem 7 (Hoeffding) *Let $z_0, \ldots z_{N-1}$ be independent random variables, of mean μ and bounded by $a \leq z_n \leq b$, define $\bar{z} = \frac{1}{N}\sum_{n=0}^{N-1} z_n$. Then for $\epsilon > 0$,*

$$\mathcal{P}\left(\bar{z} - \mu > \epsilon\right) \leq e^{\frac{-2N\epsilon^2}{(b-a)^2}} \tag{57}$$

We want to apply this inequality to the sum $r_x(\tau) = \sum_{n=0}^{N-1} z(t)$, with $z(t) = x(t)x((t+\tau)\text{mod}N)$, and $x(0) \ldots x(N-1)$ independent, identically distributed. The $z(t)$ are not independent (as required in Theorem 7), but their dependence is very slight, so the sum can be reduced to *three* sums of independent variables, as shown in the following sequence of Lemmas.

Lemma 8 *Let $\{a_1, \ldots, a_N\}$ be a permutation of $\{1, \ldots, N\}$. Then the set of ordered pairs $S = \{(1, a_1), \ldots, (N, a_N)\}$ can be partitioned into three disjoint sets S_1, S_2, S_3, of respective cardinality n_1, n_2, n_3, such that*

1. *No two pairs which fall in a single S_i have a common element of $\{1, \ldots, N\}$ (i.e., if $(n, a_n), (m, a_m) \in S_i$, $n \neq m$, then $n \neq a_m$ and $m \neq a_n$).*
2. *$n_i \geq \frac{N}{5}$, $i = 1, 2, 3$*

Proof: We perform the classification by induction. For a given n, assume that the pairs $(1,a_1),\ldots,(n,a_n)$ have been classified in disjoint sets $S_1^{(n)}$, $S_2^{(n)}$, $S_3^{(n)}$ which satisfy condition 1. Now consider a new pair $(n+1,a_{n+1})$. Since there are at most two pairs in S with an element in common with $(n+1,a_{n+1})$, at least one of the *three* $S_i^{(n)}$ will have none of these pairs and therefore condition 1 is maintained if $(n+1,a_{n+1})$ is added to it. This implies by induction that it is possible to partition S into sets S_1, S_2, S_3 satisfying condition 1.

Now consider their cardinalities n_1, n_2, n_3. Assume that $2n_i < n_j$ for some i,j. Since there are at most $2n_i$ elements in the pairs of S_i, and S_j has more *pairs*, then at least one pair in S_j shares no elements with those of S_i. Therefore this pair can be moved to S_i, maintaining condition 1. Repeating this procedure will lead to a partition S_1, S_2, S_3 satisfying condition 1 and $2n_i \geq n_j \ \forall i,j$. If n_1 is, for example, the minimum of the n_i, then $N = n_1 + n_2 + n_3 \leq n_1 + 2n_1 + 2n_1 = 5n_1$ which implies condition 2 is satisfied.

□

Lemma 9 *Let $N \geq 3$, and $x(0), x(1), \ldots, x(N-1)$ be independent identically distributed random variables. Fix $1 \leq \tau < N$. Then $r_x(\tau)$ can be expressed as $r_x(\tau) = \Sigma_1 + \Sigma_2 + \Sigma_3$, where each Σ_i is the sum of n_i independent, identically distributed random variables, and $n_i \geq \frac{N}{5}$.*

Proof: For the permutation $a_1,\ldots,a_N$ given by the circular shift $a_n = (n+\tau)\mathrm{mod}N$, perform the classification into sets S_1, S_2, S_3 of Lemma 8. Then for each i choose

$$\Sigma_i := \sum_{(n,a_n)\in S_i} x_n x_{a_n} \tag{58}$$

By construction of the sets S_i, the terms in the sum (58) are independent, identically distributed.

□

Now we return to the rest of Theorem 3.

Part 2: Assume $x(0),\ldots,x(N-1)$ are *bounded* random variables, $|x(t)| \leq K$. Pick $1 \leq \tau < N$. From Lemma 9, $r_x(\tau) = \Sigma_1 + \Sigma_2 + \Sigma_3$, where each Σ_i is the sum of n_i independent, identically distributed random variables, with zero mean and bounded in $[-K^2, K^2]$. Invoking Hoeffding's inequality and $n_i \geq \frac{N}{5}$, we have

$$\mathcal{P}\left(\frac{r_x(\tau)}{N} > \epsilon\right) \leq \sum_{i=1}^{3} \mathcal{P}\left(\frac{\Sigma_i}{n_i} > \epsilon\right) \leq \sum_{i=1}^{3} e^{\frac{-n_i\epsilon^2}{2K^4}} \leq 3e^{\frac{-N\epsilon^2}{10K^4}} \tag{59}$$

The same argument can be employed to bound $\mathcal{P}\left(-\frac{r_x(\tau)}{N} > \epsilon\right)$, for each value of τ. This implies

$$\mathcal{P}\left(\max_{1\leq\tau<N} \frac{|r_x(\tau)|}{N} > \epsilon\right) \leq 6Ne^{\frac{-N\epsilon^2}{10K^4}} = 6e^{Log(N)\left(1-\frac{N\epsilon^2}{10K^4 Log(N)}\right)} \tag{60}$$

Now choose $0 < \rho < 1$. The complement of $W_{N,\gamma,N-1}$ can be written as

$$W^C_{N,\gamma,N-1} = \left\{\max_{1\leq\tau<N} \frac{|r_x(\tau)|}{r_x(0)} > \gamma\right\} \subset \left\{\max_{1\leq\tau<N} \frac{|r_x(\tau)|}{N} > \gamma\rho\right\} \cup \left\{\frac{1}{N}\sum_{t=0}^{N-1} x(t)^2 < \rho\right\} \tag{61}$$

The probability of the first set is bounded by (60), setting $\epsilon = \gamma\rho$. The probability of the second set can be bounded by another use of the Hoeffding inequality, applied to the bounded IID random variables $x(t)^2$. Putting everything together,

$$\mathcal{P}\left(W^C_{N,\gamma,N-1}\right) \leq 6e^{Log(N)\left(1-\frac{N\gamma^2\rho^2}{10K^4 Log(N)}\right)} + e^{\frac{-2N(1-\rho)^2}{K^4}} \tag{62}$$

The second term clearly goes to to zero as $N \to \infty$, and the same happens with the first term since by hypothesis $\gamma_N \sqrt{\frac{N}{log(N)}} \xrightarrow{N\to\infty} \infty$.

Part 3: Assume $x(0), \ldots, x(N-1)$ are *Gaussian* random variables, $x(t) \sim \mathcal{N}(0,1)$. Choosing $K(N) = \sqrt{2Log(N)}$, define the random variables $v(t), t = 0, \ldots, N-1$ by truncation:

$$v(t) = \begin{cases} x(t) & if |x(t)| \leq K(N) \\ 0 & otherwise \end{cases} \tag{63}$$

$$\mathcal{P}(x \neq v) \leq N\mathcal{P}(x(t) \neq v(t)) = N\mathcal{P}(|x(t)| > K(N)) \leq \frac{NC}{K(N)}e^{\frac{-K(N)^2}{2}} = \frac{C}{\sqrt{2Log(N)}} \tag{64}$$

In (64) $x = (x(0), \ldots, x(N-1))$, $v = (v(0), \ldots, v(N-1))$, and the second inequality follows from a standard bound to the tail of the normal distribution (C is a constant). Observing that

$$\mathcal{P}(x \notin W_{N,\gamma,N-1}) \leq \mathcal{P}(v \notin W_{N,\gamma,N-1}) + \mathcal{P}(x \neq v) \tag{65}$$

it remains to show that $\mathcal{P}(v \notin W_{N,\gamma,N-1})$ also vanishes as $N \to \infty$. Since the variables $v(t)$ are bounded by $K(N)$, (62) gives

$$\mathcal{P}(v \notin W_{N,\gamma,N-1}) \leq 6e^{Log(N)\left(1-\frac{N\gamma^2\rho^2}{10K^4 Log(N)}\right)} + e^{\frac{-2N(1-\rho)^2}{K^4}} \tag{66}$$

The second term clearly has limit 0 as $N \to \infty$. The first term also goes to 0, since by hypothesis $\frac{N\gamma^2\rho^2}{K^4 Log(N)} = \frac{\rho^2}{4}\frac{N\gamma^2}{Log(N)^3}$ goes to infinity.

Remarks on the Proof of Theorem 5

The fact that a uniform bound is being applied to the cumulative periodogram means that we are imposing a number of constraints of the order of the sample size, as in Theorem 3 parts 2, 3; this again precludes simple arguments based on averaging.

The key observation, which led Bartlett (see [2]) to propose this test, is to notice that the stochastic properties of the cumulative periodogram are similar to those used for tests on empirical distribution functions. The maximum deviation between an empirical distribution and the true distribution function forms the basis of the Kolmogorov-Smirnov test (see [4]), which has well known asymptotic properties. The connection with the cumulative periodogram can be seen as follows: in the case of Gaussian white noise, the periodogram values are independent and exponentially distributed (see [5]), which implies ([4], Prop. 13.15) that the normalized cumulative periodogram values $\frac{S_m}{S_N} = \frac{1}{N\|x\|^2}\sum_{k=0}^{m-1} s_x(k)$ have the same joint distribution as an ordered sample of uniform $(0,1)$ variables. From these arguments it follows that

$$\sqrt{N} \sup_{1\leq m<N} \left| \frac{1}{N\|x\|^2}\sum_{k=0}^{m-1} s_x(k) - \frac{m}{N} \right|$$

converges in law to a fixed distribution. Since $\frac{1}{\eta_N\sqrt{N}} \stackrel{N\to\infty}{\longrightarrow} 0$, then

$$\frac{1}{\eta_N} \sup_{1\leq m<N} \left| \frac{1}{N\|x\|^2}\sum_{k=0}^{m-1} s_x(k) - \frac{m}{N} \right| \stackrel{\mathcal{P}}{\longrightarrow} 0$$

which proves the theorem.

An additional remark is that although this proof is valid for Gaussian noise, there is indication in [2] that the asymptotic properties are insensitive to the noise distribution.

Proof of Proposition 6

For a fixed $\tau \neq 0$, referring to [4] (proposition 6.31), we find that the random process $z(t) = x(t)x(t+\tau)$ is ergodic, so with probability 1,

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{t=-N}^{N} x(t+\tau)x(t) = E[x(t+\tau)x(t)] = 0 \tag{67}$$

Therefore $W_{0,\infty}$ has probability 1 (countable intersection of probability 1 sets).

Acknowledgements

The author would like to thank John Doyle for motivation and helpful discussions at Caltech, Geir Dullerud for useful suggestions which helped simplify the proofs, and Stefano Soatto, Giorgio Picci and Adelchi Azzalini for useful references. This work was supported by AFOSR, NSF and by the Universidad de la Republica, Uruguay.

References

[1] Anderson R.L., Anderson T.W., *Distribution of the circular serial coefficient for residuals from a fitted Fourier series*, Ann. Math. Statist., 21, pp. 59-81, 1956.

[2] Bartlett M.S., *An Introduction to Stochastic Processes*, Cambridge University Press, 1955.

[3] Boyd, S.P., L El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadelphia, Pennsylvania, 1994.

[4] Breiman L., *Probability*, Addison-Wesley, 1968.

[5] Brillinger D., *Time Series Data Analysis and Theory*, Mc Graw-Hill, 1981.

[6] Dahleh M.A, Theodosopoulos T.V., Tsitsiklis J., *The sample complexity of worst-case identification of FIR linear systems*, Systems and Control Letters, 20, pp 157-166, 1993.

[7] Feron E., *Analysis of Robust $\mathcal{H}_2$ Performance with Multipliers*, Proc. 1994 CDC, pp. 2015-2020.

[8] Jenkins G., Watts D., *Spectral Analysis and its Applications*, Holden-Day, 1968.

[9] Helmicki A.J., Jacobson C.A., Nett C. N, *Control Oriented System Identification: A Worst-Case/Deterministic Approach in $\mathcal{H}_\infty$*, IEEE Trans. A. C., Vol 36 (10), pp. 1163-1176, 1991.

[10] Hoeffding W., *Probability Inequalities for Sums of Bounded Random Variables*, American Statistical Association Journal, 59, pp. 13-30, March 1963.

[11] Ljung, L., *System Identification Theory for the User*, Prentice-Hall, 1987.

[12] Mari J., *A counterexample in Power Signals Space*, submitted to IEEE Trans A.C., 1994.

[13] Megretski A., *S-Procedure in Optimal Non-Stochastic Filtering,* Tech. Rep. Code Trita/Mat-92-0015, Royal Inst. of Tech., Sweden, 1992.

[14] Megretski A., Treil S., *Power Distribution Inequalities in Optimization and Robustness of Uncertain Systems*, Journal of Math. Sys., Est. & Control, Vol 3, No.3, pp 301-319, 1993.

[15] Packard A., Doyle J.C., *The Complex Structured Singular Value,* Automatica, Vol. 29, No. 1, pp. 71-109, 1993.

[16] Paganini F., *Set Descriptions of White Noise and Worst Case Induced Norms,* Proceedings 1993 CDC, San Antonio, Texas, pp. 3658-3663.

[17] Paganini F., D'Andrea R., and Doyle J., *Behavioral Approach to Robustness Analysis* Proceedings 1994 ACC, Baltimore, MD, pp. 2782-2786.

[18] Paganini F., Doyle J., *Analysis of Implicit Uncertain Systems, Parts I and II,* CDS Tech. Rep. CDS94-018, California Institute of Technology, 1994, submitted to IEEE Trans A.C.

[19] Paganini F., *Necessary and Sufficient Conditions for Robust $\mathcal{H}_2$ Performance*, submitted to 1995 CDC.

[20] Peters M.A., and Stoorvogel A.A., *Mixed $\mathcal{H}_2/\mathcal{H}_\infty$ Control in a Stochastic Framework*, Lin. Alg. & Apps., Vol 205-206, pp. 971-996, 1994.

[21] Poolla K., Tikku A., *On the Time Complexity of Worst-Case System Identification*, IEEE Trans AC., Vol 39 (5), pp. 944-950, 1994.

[22] Rudin W., *Principles of Mathematical Analysis*, Mc Graw Hill, New York 1952.

[23] Ruelle D., *Chaotic Evolution and Strange Attractors,* Cambridge University Press, 1989.

[24] Shamma, J., *Robust Stability with Time Varying Structured Uncertainty*, IEEE Trans. A. C., Vol 39, 4, pp 714-724, 1994.

[25] Stoorvogel A.A., *The Robust $\mathcal{H}_2$ Control Problem: A Worst-Case Design*, IEEE Trans. A.C., Vol 38, 9,pp. 1358-1370, 1993.

[26] Tse D., Dahleh M.A., Tsitsiklis J., *Optimal Asymptotic Identification Under Bounded Disturbances*, IEEE Trans A.C., Vol 38 (8), pp. 1176-1190, 1993.

[27] Venkatesh, S. R., Dahleh M.A., *Classical system identification in a deterministic setting.* Preprint, submitted to 1995 CDC.

[28] Wiener N., *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Wiley, New York, 1950.

[29] Yakubovich, V. A., *S-procedure in nonlinear control theory*, Vestnik Leningrad University, pp. 62-77, 1971 (English translation in Vestnik Leningrad Univ. 4:73-93, 1977).

[30] Zhou K., Glover K., Bodenheimer B., Doyle J., *Mixed $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Performance Objectives I: Robust Performance Analysis*, IEEE Trans. AC., Vol 39, 8, pp. 1564-1574, 1994